PERCEPTUAL BOUNDARY BETWEEN A SINGLE AND A GEMINATE STOP IN JAPANESE

Shigeaki Amano, Ryoko Mugitani, and Tessei Kobayashi

NTT Communication Science Laboratories, NTT Corporation,
amano@cslab.kecl.ntt.co.jp

ABSTRACT

Hirata and Whiton [J. Acou. Soc. Am. 118(3), 1647-1660, 2005] revealed that the production boundary between a single and a geminate stop in Japanese is invariant over speaking rates in terms of the ratio of stop closure duration to word duration (closure-word ratio). The present study addressed the question of whether the ratio is also invariant for the perceptual boundary. An experiment was conducted to obtain the perceptual boundary between a single and a geminate stop at slow, normal and fast speaking rates. The results showed that the closure-word ratio at the perceptual boundary between a single and a geminate stop did not coincide with that at its production boundary as obtained by Hirata and Whiton. However, the closure-word ratio was consistent within each stimulus item for all speaking rates, although it was different among the stimulus items. The results suggest that the closure-word ratio at the perceptual boundary between a single and a geminate stop is invariant over speaking rates within an item, but some item-related factors affect the perceptual boundary.

Keywords: geminate stop, perception, production

1. INTRODUCTION

A geminate stop (/Q/) (cf. [9], p.13) has a longer closure duration than a single stop in Japanese. There have been many studies on the perception of the geminate stop.

For example, Fukui [1] found that the closure duration is a primary cue for the perception of a geminate stop. He also suggested that a secondary cue is the duration of a vowel that precedes a single or geminate stop (hereafter referred to as a preceding vowel).

Watanabe and Hirato [10] revealed that the preceding vowel duration affects the closure duration at the perceptual boundary between a single and a geminate stop. They found a linear relationship between the preceding vowel duration and closure duration at the perceptual boundary. That is, if the preceding vowel duration is long, the closure duration at the perceptual boundary becomes long. In a follow-up study, Hirato and Watanabe [4] found that the following vowel duration has no effect on the closure duration at the perceptual boundary between a single and a geminate stop.

In addition to these perception studies, there have been geminate stop production studies. For example, Hirata and Whiton [3] analyzed a single and a geminate stop in 2-mora Japanese words and nonwords pronounced by four Japanese at slow, normal, and fast speaking rates. They measured several parameters related to single and geminate stops such as closure duration, preceding vowel duration, voice onset time, word duration where a single or a geminate stop is involved, the ratio of closure duration to preceding vowel duration (hereafter referred to as the closure-vowel ratio), and the ratio of closure duration to word duration (hereafter referred to as the closure-word ratio).

They found that the closure-vowel ratio is the second best parameter and the closure-word ratio is the most powerful parameter with which to distinguish a single stop from a geminate stop. That is, the production boundary between a single and a geminate stop is constant in terms of the closure-word ratio irrespective of speaking rate. The value of the closure-word ratio was 0.35.

Because speech perception and speech production are closely related, the closure-vowel ratio and the closure-word ratio, which are invariant parameters in the production of a geminate stop, might also be invariant parameters in its perception. Based on this idea, the present study addresses the question of whether the closure-vowel ratio or the closure-word ratio at a perceptual boundary coincides with that at the production boundary obtained by Hirata and Whiton [3].

2. METHOD

2.1. Participants

Forty Japanese adults (20 males and 20 females) participated in the experiments. Their mean age was 25.4 (SD = 0.74, Min = 20, Max = 39). They exhibited no symptoms of hearing disability. They were paid for their participation.

2.2. Stimulus

2.2.1. Original materials

The original materials were five 2-mora items (/bipa/, /guku/, /kuku/, /kuto/, /tapi/). A professional female Japanese narrator pronounced the items with an HL-type pitch accent (i.e., the first mora had a high pitch and the second mora had a low pitch) at a normal speaking rate without devoicing the vowels. These pronounced materials were digitally recorded with a 16-bit quantization and a 44.1-kHz sampling frequency, and stored in a computer as speech files. Table 1 shows the duration of the portions of the original materials

Table 1: Duration of original materials (ms).

	First mora			Second mora	
Item	Burst	Vowel	Closure	Burst	Vowel
/bipa/	44	97	161	11	222
/guku/	52	117	120	37	214
/kuku/	35	92	114	41	193
/kuto/	36	102	143	17	233
/tapi/	28	109	166	11	197

2.2.2. Stimulus set

A stimulus set was made from the original materials by controlling the duration of the preceding vowel and the closure duration of an unvoiced plosive between the first and second mora.

The duration of the preceding vowel was controlled with the STRAIGHT method [6], which can synthesize high-quality time-stretched speech. The time-stretch rates were 50% (i.e., reduced to half duration) or 150% (i.e., increased to one and a half duration). In addition to these rates, a 100% time-stretch rate (i.e., original duration) was used. This operation corresponded to controlling the speaking rate of the stimulus, because the durational change in vowels is much greater than that in consonants in Japanese when the speaking rate is changed [2], and because the preceding vowel duration is an important factor as regards the perception of a geminate stop [10].

To create stimulus continua from a single stop to a geminate stop, the closure duration of an unvoiced plosive was controlled by inserting a silence or removing part of a closure in 10-ms steps. The range of this durational change was fixed at 200 ms, but the end points were changed according to the time-stretch rate of the preceding vowel. The end points were -60 ms and 140 ms for a 50% time-stretch rate, -20 ms and 180 ms for a 100% time-stretch rate, and 20 ms and 220 ms for a 150% time-stretch rate. There were 315 stimuli (5 items x 3 time-stretch rates x 21 closure durations) in the test stimulus set.

2.3. Procedure

The stimuli were stored in a personal computer and they were diotically presented to participants through a D/A converter and headphones at a comfortable presentation level. The stimulus presentation order was randomized for each participant. Two response buttons were displayed at the center of a computer screen. The left button was for an item with a single stop (e.g., /bipa/), and the right button was for an item with a geminate stop (e.g., /biQpa/). These assignments were indicated on the response buttons in hiragana (Japanese phonetic) characters, which corresponded to the stimulus presented through the headphones in each trial.

The participants were instructed to select one of the two response buttons, one for a single stop the other for a geminate stop, by a mouse click according to what they heard. The participant's responses were automatically saved in a personal computer. After the selection, they were asked to click the ``next" button shown below the two response buttons to receive a new stimulus. The participants received 1575 test stimuli (315 stimuli x 5 repetitions), and took a short break after every 315 trials.

2.4. Results

The response rate for a geminate stop was calculated by dividing the number of participants who provided a geminate stop response by the total number of participants. The response rate was obtained at every 10-ms point of the closure duration for each item at each time-stretch rate. A logistic curve was fitted to the response rate as a function of the closure duration for each item at each time-stretch rate.

The perceptual boundary between a single and a geminate stop was obtained as the closure duration that gave a 0.5 response rate on the fitted logistic curve for each item and time-stretch rate. The closure-vowel ratio and closure-word ratio were calculated by dividing the closure duration at the perceptual boundary by the preceding vowel duration and by the word duration, respectively (Table 2).

The closure-vowel ratios were inconsistent among time-stretch rates and among items. They were different from the closure-vowel ratio (1.52 or 1.53) which Hirata and Whiton [3] obtained. This means that the closure-vowel ratio for perception does not coincide that for production and that it is not invariant across time-stretch rates.

The mean of the closure-word ratio at the perceptual boundary is significantly different from 0.35 [$t(15) = 5.34$, $p < .0001$, two-tailed]. This result indicates that the closure-word ratio at the perceptual boundary does not coincide with that at the production boundary, which is 0.35 in Hirata and Whiton's study [3].

Analysis of variance was conducted for the closure-word ratio with the two factors of time-stretch rate and item. The time-stretch rate factor was not significant. However, the item factor was significant [$F(4,8) = 72.7$, $p < .0001$]. Paired comparisons were significant at the 5% level between /bipa/-/guku/, /bipa/-/tapi/, /guku/-/kuku/, /guku/-/kuto/, /guku/-/tapi/, /kuku/-/kuto/, and /kuku/-/tapi/. These results indicate that the closure-word ratio is different among items but not among time-stretch rates.

3. DISCUSSION

The results of this study indicate that the closure-vowel ratio and the closure-word ratio of the perceptual boundary between a single and a geminate stop do not coincide with that of a production boundary. This perception-production discrepancy in phoneme boundary is also found in previous studies (e.g., [7]).

However, the closure-word ratio at the perceptual boundary is very consistent across the time-stretch rate in this study (Table 2). This invariance in the time-stretch rate is compatible with the results reported by Hirata and Whiton (2005), namely that the closure-word ratio at a production boundary is unaffected by the speaking rate. The results of this study suggest that the closure-word ratio is a robust parameter against speaking rate in the perception of a geminate stop as well as in its production.

Table 2: Closure-vowel ratio and closure-word ratio for each item and time-stretch rate.

Item	Time-stretch rate of preceding vowel duration (%)	Closure duration at perceptual boundary (ms)	Preceding vowel duration (ms)	Word duration (ms)	Closure-vowel ratio	Closure-word ratio
/bipa/	50	191	48	487	3.94	0.39
	100	214	97	536	2.21	0.40
	150	238	145	584	1.64	0.41
/guku/	50	166	58	481	2.84	0.34
	100	188	117	540	1.61	0.35
	150	200	175	598	1.14	0.33
/kuku/	50	168	46	429	3.65	0.39
	100	186	92	475	2.02	0.39
	150	203	138	521	1.47	0.39
/kuto/	50	201	51	480	3.94	0.42
	100	222	102	531	2.18	0.42
	150	237	153	582	1.55	0.41
/tapi/	50	194	54	455	3.56	0.43
	100	224	109	510	2.06	0.44
	150	243	163	564	1.49	0.43

It is very interesting that the same parameter, namely the closure-word ratio, is robust against the speaking rate both in perception and production. However, this robustness of speaking rate in perception is only found within an item. The closure-word ratio at a perceptual boundary is different across items (Table 2).

What causes this item-related discrepancy in the closure-word ratio at a perceptual boundary? There are several possibilities. One is the vowel onset timing, which is related to a power increase over a short time. Kato, Tsuzaki, and Sagisaka [5] showed that this vowel onset timing affects speaking rate perception. Because the vowel onset timing depends on the phonemic context it might explain the item-related discrepancy.

Another possibility is a spectral cue including a formant transition. The formant transition differs according to the sequence of consonants and vowels. Because each item has its own sequence of consonants and vowels, this kind of spectral cue is a candidate cause for the item-related discrepancy.

In addition, prosodic cues such as pitch transition might be a cause of the item-related discrepancy. Although the pitch accent of the stimuli was fixed as HL-type in this study, the pitch transition is not exactly the same for all items. Such differences in pitch transition might affect the perception of a single/geminate stop.

Although this study provides clear results in relation to the closure-word ratio at the perceptual boundary between a single and a geminate stop, we should be aware of certain limits in terms of generalization.

Firstly, stimuli are presented without a carrier sentence in this study whereas they are embedded in a carrier sentence in Hirata and Whiton’s work (2005). It is possible that the item-related discrepancy disappears when the stimuli are embedded in a carrier sentence, because a sentence would have many more kinds of timing cue than a single word. Further study is necessary using stimuli with a carrier sentence.

Secondly, the stimulus duration was changed only in the preceding vowel and closure in this study. That is, the stimulus duration was fixed in other segments such as following vowels. This fixation might have led to the presence of unknown artifacts in the current results. Although the preceding vowel and closure are the main cues for the perception of a geminate stop [1], other segments might have an effect on the perception. Recently, Ofuka, Mori, and Kiritani [8] suggested that the following-vowel duration has an effect on the perceptual boundary between a single and a geminate stop. The effects of the segment duration including the following vowel should be explored in further research.

Finally, the results are based on items with little variety. For example, only a geminate stop with a plosive was used in this study. But a geminate stop can occur with a fricative or an affricate. Therefore, the perceptual boundary between a single and a geminate stop should be investigated with fricatives and affricates to allow more general conclusions to be drawn.

4. REFERENCES

[1] Fukui, S. 1978. Perception for the Japanese stop consonants with reduced and extended durations. *Onsei Gakkai Kaiho* 159, 9-12. (In Japanese).

[2] Hiki, S., Kanamori, Y., Oizumi, J. 1967. On the duration of phoneme in running speech. *Denki-Tsushin-Gakkai-Zasshi* 50, 69-76. (In Japanese).

[3] Hirata, Y., Whiton, J. 2005. Effects of speaking rate on the single/geminate stop distinction in Japanese. *The Journal of the Acoustical Society of America* 118, 1647-1660.

[4] Hirato, N. and Watanabe, S. 1987. The relation between the perceptual boundary of voiceless pslosives and their moraic counterparts and the length of vowels following closure. *Onsei-Gengo* 2, 99-106. (In Japanese).

[5] Kato, H., Tsuzaki, M., Sagisaka, Y. 2003. Functional differences between vowel onsets and offsets in temporal perception of speech: Local-change detection and speaking-rate discrimination. *The Journal of the Acoustical Society of America* 113, 3379-3389.

[6] Kawahara, H., Masuda-Katsuse, I., de Cheveigne, A. 1999. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication* 27, 187-207.

[7] Miller, J. L., Green, K. P., Reeves, A. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for voicing contrast. *Phonetica* 43, 106-115.

[8] Ofuka, E. 2003. Perception of a Japanese geminate stop /tt/: the effect of pitch type and acoustic characteristics of preceding/following vowels. *Journal of the phonetic society of Japan* 7, 70-76, (In Japanese).

[9] Toda, T. 2003. *Second language speech perception and production.* Lanham: University Press of America.

[10] Watanabe, S., Hirato, N. 1985. The relation between the perceptual boundary of voiceless plosives and their moraic counterparts and the duration of the previous vowel. *Onsei-Gengo* 1, 1-8. (In Japanese).